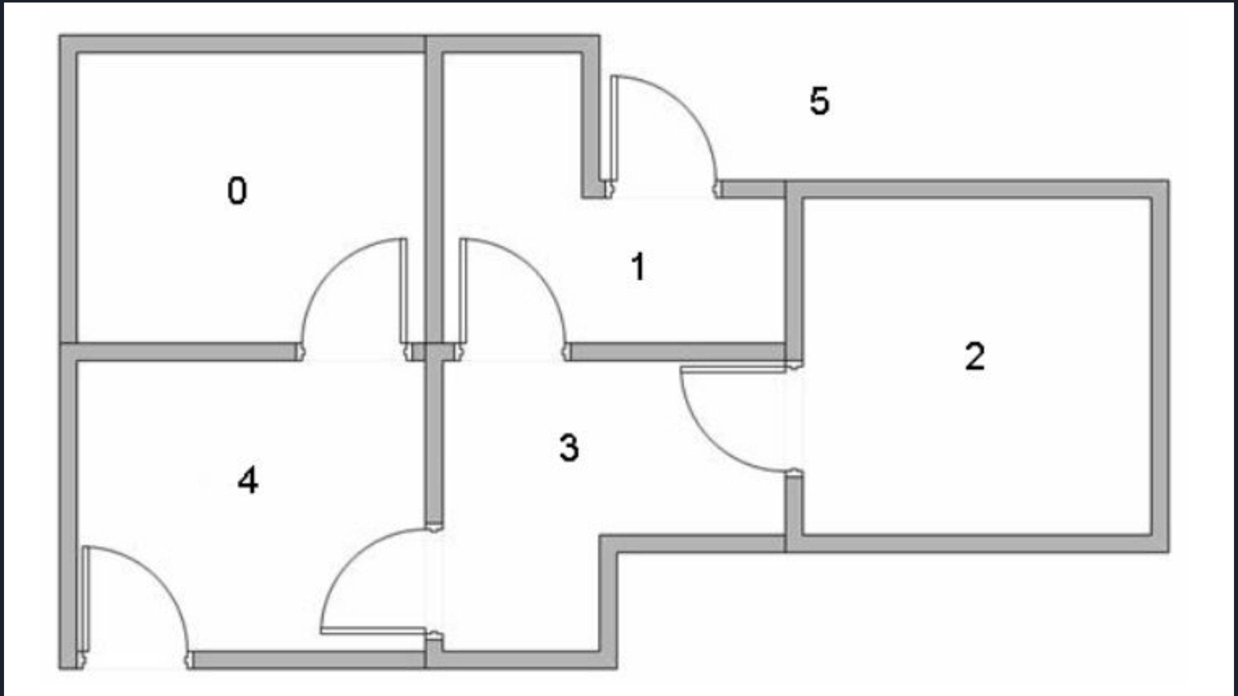
Q-Learning to solve a maze problem

Joseph Gonzalez
Oscar Guarnizo
Andres Banda
Fabricio Crespo

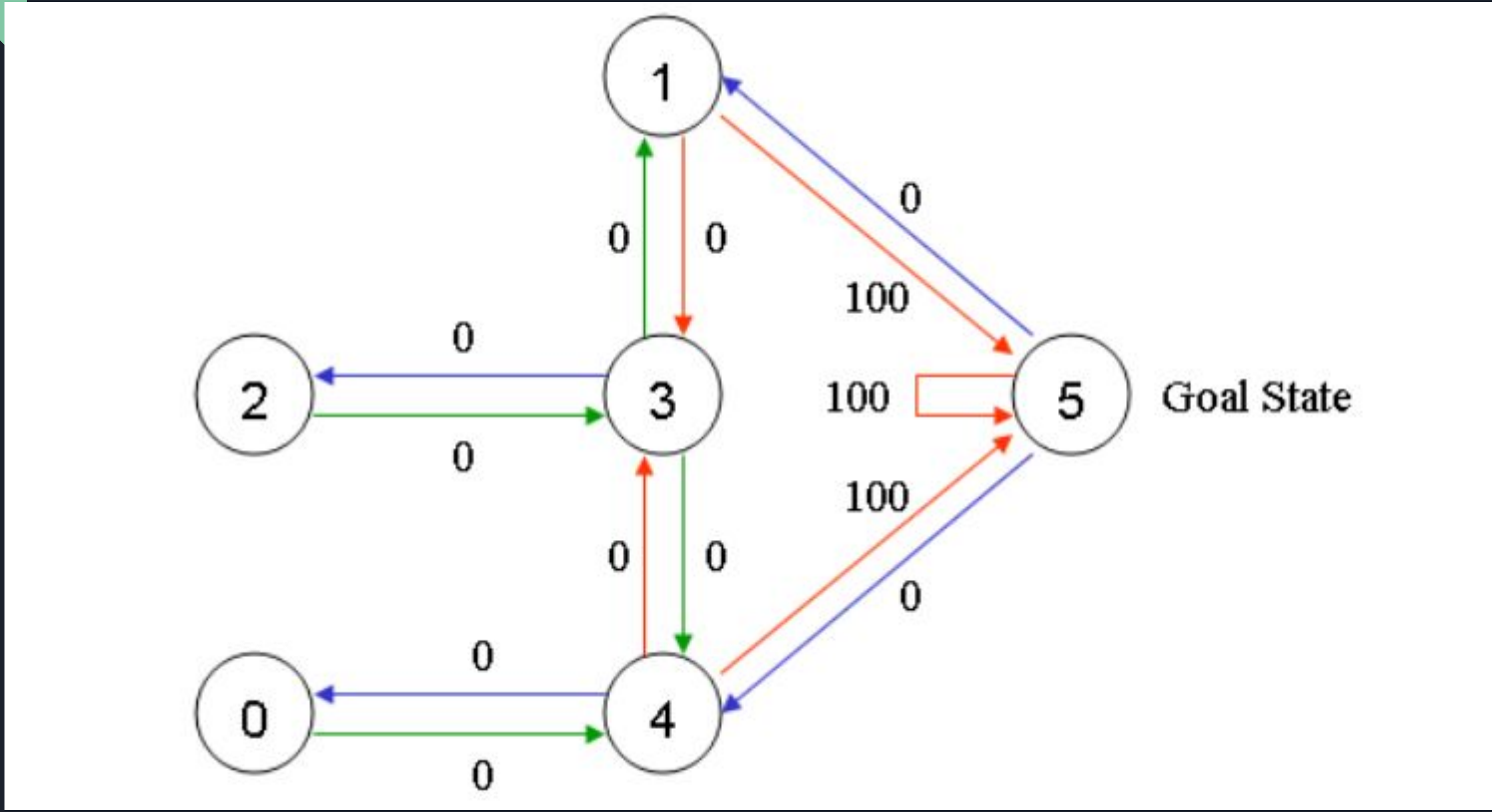
Outline

1. Introduction
2. Model Proposal
3. Results
4. Conclusion

Introduction: Previous Work

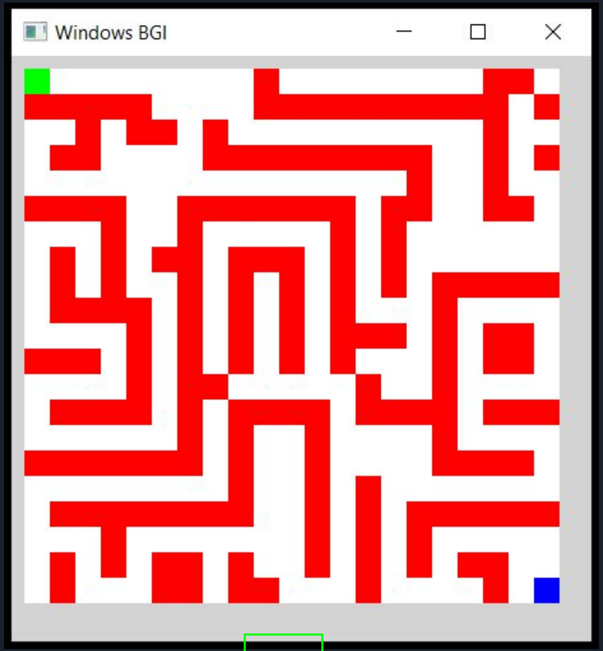


Introduction: Previous Work



Model Proposal

- States: equal to the dimension of the maze, m x n.
- Actions: Left, Top, Right, Down



n

m

Q-matrix

Actions :4

$$Q = \begin{bmatrix} -500 & -500 & 0 & -500 \\ 0 & -500 & 0 & -500 \\ \cdots & \cdots & \cdots & \cdots \\ -500 & 0 & 0 & -500 \\ 0 & 0 & -500 & -500 \end{bmatrix}$$

States: n*m = 20*20 = 400

R-matrix

Actions :4

$$R = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 100 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

States:

n*m = 20*20 = 400

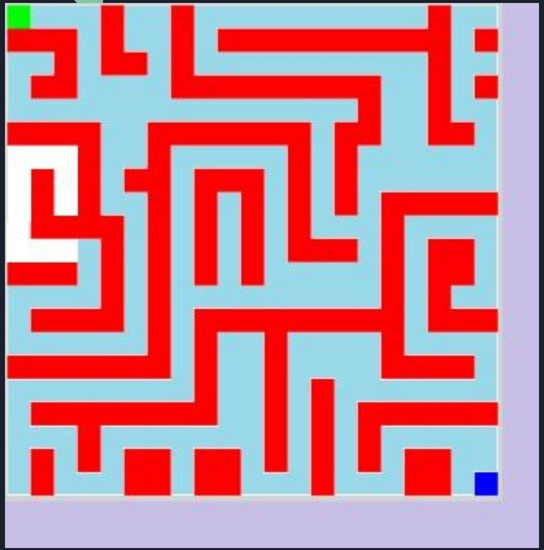
We just give a reward of 100 when ant agent performs an action which moving it to the final position.

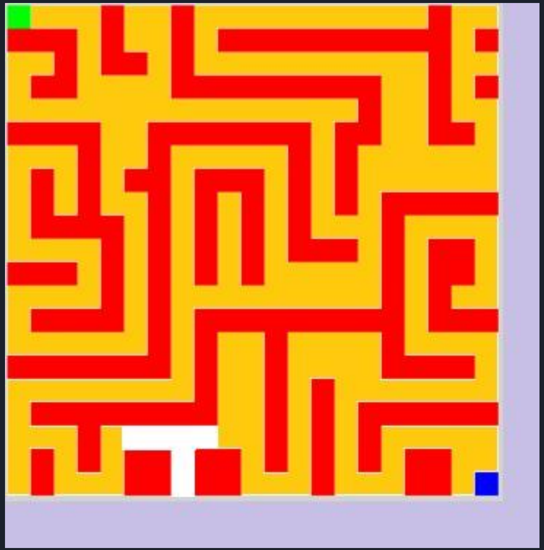
Episodes and Hyperparameters

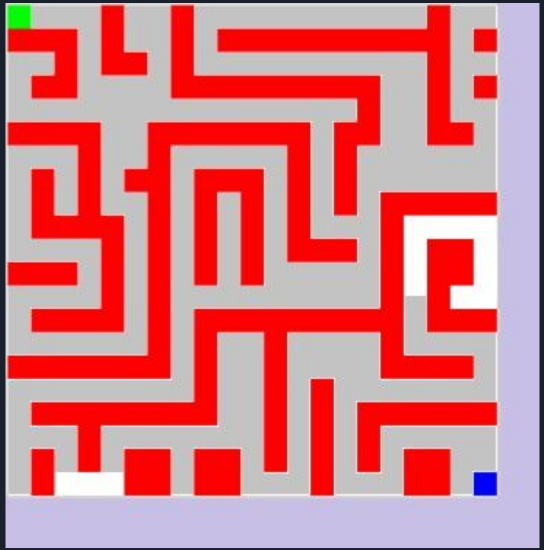
- An episode is defined by a number of iteration: 200
- Gamma: it is used to the rule of learning in Q-learning.
- Epsilon: it is used to defined the influences of exploration over the exploitation techniques.

Results

Three different training with 400 states and 4 possible actions:





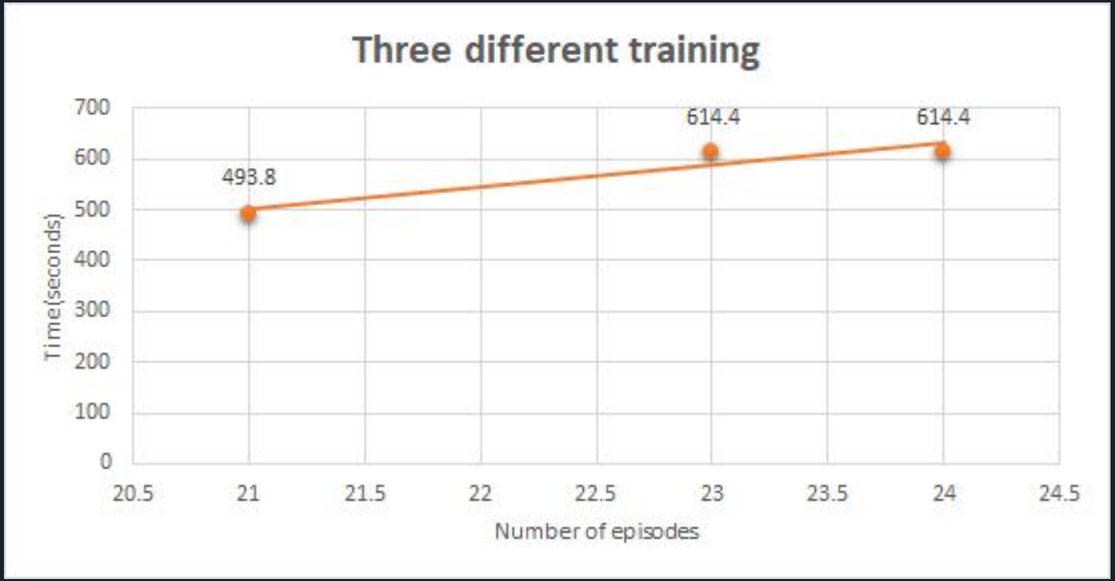


Total episodes: 21
Time: 8.24 minutes
Labyrinth completed: 96.75%

Total episodes: 23
Time: 10.24 minutes
Labyrinth completed: 98.5%

Total episodes: 24
Time: 10.42 minutes
Labyrinth completed: 96.75%

Comparison of different training.



Results

Training Time Average:

	The time that the agent takes to training on the Labyrinth			
	Time with 3 different training (seconds)			Average
Dimension	1st	2nd	3nd	
20x20	493.8	614.4	614.4	574.2

574.2 seconds = 9.57 minutes

Results

Training Time Average:

	The time that the agent takes to training on the Labyrinth			
	Time with 3 different training (seconds)			Average
Dimension	1st	2nd	3nd	
10x10	306	288	330	308

308 seconds = 5.13 minutes

Conclusiones

- **Q-learning** provides **agents** with the capability of **learning** to act **optimally** figuring out the maze solution.

- **States(400)** and **actions(4)** determine how this **learning** process is achieved.

Conclusiones

- Q-learning **optimization** is done through **rewards** received (R matrix) .

- Q-learning Maze **implementation** demonstrate to be **efficient.**

Conclusiones

- A minimum of 96% of the labyrinth completed explored by the agent in 10 minutes.
- The time required to train the algorithm with respect to the maze dimension is **linear**.
- If the dimension of the labyrinth decreases, the agent takes less time to train.

THANK YOU

Git Hub Repository: https://github.com/ZosoV/QLearningMazeSolving.git

Joseph González: joseph.gonzalez@yachaytech.edu.ec

Oscar Guarnizo: oscar.guarnizo@yachaytech.edu.ec

Andrés Banda: jorge.banda@yachaytech.edu.ec

Fabricio Crespo: jonnathan.crespo@yachaytech.edu.ec